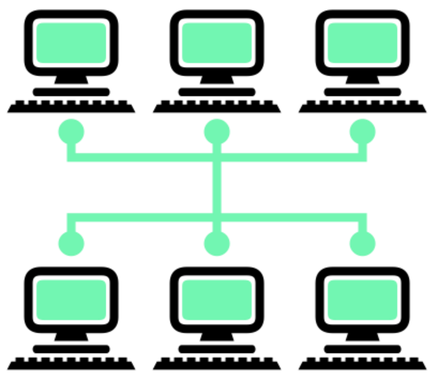
molgenis

Enabling FAIR data analysis
The sandbox and RD3 sample database

In the rare disease community within various projects researchers are joining forces to solve rare diseases. Omics data of patient samples is gathered from centers all over the continent, or beyond. Such analysis requires FAIR data. To ensure accessibility and enable researchers from different centers to jointly analyze these samples we have created the analytical sandbox/RD3 tandem, that consists of the sandbox, an HPC compute environment, and to aid findability and reusability, the RD3 (Rare Disease Data about Data) database. Metadata on subjects, samples and laboratory experiments are stored in RD3.

Virtual Research Environment

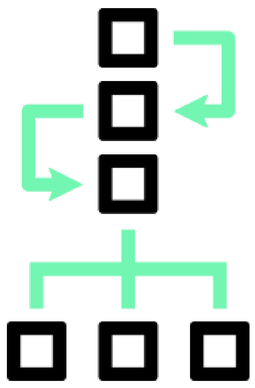
We can reproduce a complete virtual 'HPC' cluster with a single command using OpenStack and Ansible, for example at Groningen Center of Information Technology "Shikra" cloud or at the "Embassy" cloud of the European Bioinformatics Institute.



Created by Trevor Dsouza from the Noun Project

Reproducible Deployement

All software and reference data are automatically deployed using EasyBuild and versionable using Linux module system to ensure to quickly and exactly replicate analyses on multiple sites enabling QC/test, fail-over and scientific reproducibility.



Created by ArafatUddin from the Noun Project

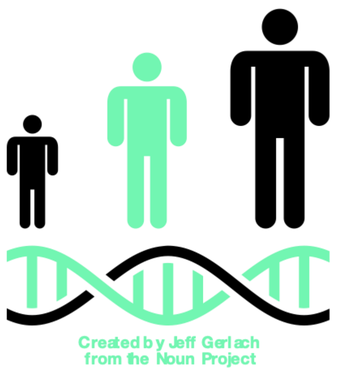
FAIR GENOMES data tracker

We created MOLGENIS 'rare and common disease data about data' (RD3) database, which enables central tracking of studies, subjects, samples, files, workflows, tasks, runs, jobs and progress monitoring of all analysis steps involved.

Created by ProSymbols from the Noun Project

Declarative Omics Pipelines

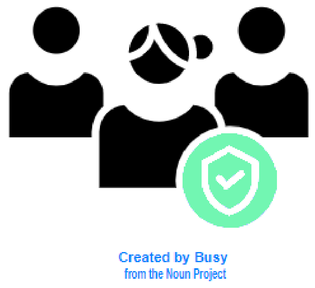
We implemented DNA, RNA and microbiome pipelines using MOLGENIS/compute workflow framework which enables fast declarative definition and logging of advanced Linux-based bioinformatics workflows.
In addition, other omics analysis tools are available for custom analysis.



Created by Jeff Gerlach from the Noun Project

Findable and Accessible

RD3 and sandbox connect to the European Genome-Phenome Archive (EGA) through Filesystem in Userspace (FUSE) and download client. Persistent Identifiers ensure unique file identification. Using Elixir AAI the data can be securely accessed.



Created by Busy from the Noun Project

Interoperable and Reusable

RD3 and sandbox use standard file formats such as VCF & PED and are early adopters of phenopackets, a GA4GH suggested standard.

Created by Vectors Market from the Noun Project

PARTNERS & SPONSORS

Authors: Lennart Johansson[1], Gerben van der Vries[1], Pieter Neerincx[1], Fernanda de Andrade[1], Dieuwke Roelofs-Prins[1], Roan Kanninga[1], Marieke Bijlsma[1], Joeri van der Velde[1], Dennis Hendriksen[1], Egon Rijpkema[5], Wim Nap[5], Henk-Jan Zilverberg[5], Fokke Dijkstra[5], Jan-Jurjen Uitterdijk[5], Gert-Jan van de Geijn[1], Steven Laurie[3], Davide Piscia[3], Giselle Kerry[4], Anand Mohan[4], MOLGENIS team[1], Dylan Spalding[4], Sergi Beltran[3], Anthony Brookes[2], Morris Swertz[1]
Contact: l.johansson@umcg.nl

1 umcg
2 University of Leicester

MOLGENIS is a fully customizable tool suite for complex life science data and computations.

MOLGENIS is open source and free.

More information: https://www.molgenis.org

3 cnag centre nacional d'anàlisi genòmica centro nacional de análisis genómico; CRG Centre for Genomic Regulation
4 EMBL-EBI
5 rijksuniversiteit groningen